Distinguishing a truncated random permutation from a random function

Shoni Gilboa*   Shay Gueron [†‡]

July 29, 2015

Abstract

An oracle chooses a function f from the set of n bits strings to itself, which is either a randomly chosen permutation or a randomly chosen function. When queried by an n-bit string w, the oracle computes $f(w)$, truncates the m last bits, and returns only the first $n-m$ bits of $f(w)$. How many queries does a querying adversary need to submit in order to distinguish the truncated permutation from a random function?

In 1998, Hall et al. [2] showed an algorithm for determining (with high probability) whether or not f is a permutation, using $O(2^{\frac{m+n}{2}})$ queries. They also showed that if $m < n/7$, a smaller number of queries will not suffice. For $m > n/7$, their method gives a weaker bound. In this manuscript, we show how a modification of the method used by Hall et al. can solve the porblem completely. It extends the result to essentially every m, showing that $\Omega(2^{\frac{m+n}{2}})$ queries are needed to get a non-negligible distinguishing advantage. We recently became aware that a better bound for the distinguishing advantage, for every $m < n$, follows from a result of Stam [3] published, in a different context, already in 1978.

Keywords: Pseudo random permutations, pseudo random functions, advantage.

1   Introduction

Distinguishing a randomly chosen permutation from a random function is a combinatorial problem which is fundamental in cryptology. A few examples where this problem plays an important role are the security analysis of block ciphers, hash and MAC schemes.

One formulation of this problem is the following. An oracle chooses a function $f : \{0,1\}^n \to \{0,1\}^n$, which is either a randomly (uniformly) chosen permutation of $\{0,1\}^n$, or a randomly (uniformly) chosen function from $\{0,1\}^n$ to $\{0,1\}^n$. An adversary selects a "querying and guessing" algorithm. He first uses it to submit q (adaptive) queries to the oracle, and the oracle responds with $f(w)$ to the query $w \in \{0,1\}^n$. After collecting the q responses, the adversary uses his algorithm to guess whether or not f is a permutation. The quality of such an algorithm (in the cryptographic context) is the ability to distinguish between the two cases (rather than successfully guessing which one it is). It is measured by the difference between the probability that the algorithm outputs a certain answer, given that the oracle chose a permutation, and the probability that the algorithm outputs the same answer, given that the oracle chose a function. This difference is called

---

*Mathematics Department, The Open University of Israel, Raanana 43107, Israel. Email: tipshoni@gmail.com

[†]Mathematics Department, University of Haifa, Haifa 31905, Israel

[‡]Intel Corporation

the "advantage" of the algorithm. We are interested in estimating Adv, which is the maximal advantage of the adversary, over all possible algorithms, as a function of a budget of q queries.

The well known answer to this problem is based on the simple "collision test" and the Birthday Problem:

$$Adv = 1 - \left(1 - \frac{1}{2^n}\right)\left(1 - \frac{2}{2^n}\right) \cdots \left(1 - \frac{q-1}{2^n}\right).$$

Since for every $1 \le k \le q - 1$

$$1 - \frac{q}{2^n} \le \left(1 - \frac{k}{2^n}\right)\left(1 - \frac{q-k}{2^n}\right) \le \left(1 - \frac{q}{2^{n+1}}\right)^2,$$

we get, for $q \le 2^n$, that

$$1 - e^{-\frac{q(q-1)}{2^{n+1}}} \le 1 - \left(1 - \frac{q}{2^{n+1}}\right)^{q-1} \le Adv \le 1 - \left(1 - \frac{q}{2^n}\right)^{\frac{q-1}{2}} \le \frac{q(q-1)}{2^{n+1}}. \tag{1}$$

This result implies that the number of queries required to distinguish a random permutation from a random function, with success probability significantly larger than, say, $1/2$, is $\Theta(2^{n/2})$.

We now consider the following generalization of this problem:

Problem: Let $0 \le m < n$ be integers. An oracle chooses $c \in \{0, 1\}$. If $c = 1$, it picks a permutation p of $\{0, 1\}^n$ uniformly at random, and if $c = 0$, it picks a function $f : \{0, 1\}^n \to \{0, 1\}^n$ uniformly at random. An adversary is allowed to submit queries $w \in \{0, 1\}^n$ to the oracle. The oracle computes $\alpha = p(w)$ (if $c = 1$) or $\alpha = f(w)$ (if $c = 0$), truncates (with no loss of generality) the last m bits from α, and replies with the remaining $(n - m)$ bits. The adversary has a budget of q (adaptive) queries, and after exhausting this budget, is expected to guess c. *How many queries does the adversary need in order to gain non-negligible advantage?*
Specifically, we seek $q_{1/2} = \min\{q \mid Adv \ge 1/2\}$ as a function of m and n.

This problem was studied by Hall et al. [2] in 1998. The authors showed (for every m) an algorithm that gives a non-negligible distinguishing advantage using $q = O(2^{(n+m)/2})$ queries. They also proved the following upper bound:

$$Adv \le 5 \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{\frac{2}{3}} + \frac{1}{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^3 \frac{1}{2^{\frac{n-7m}{2}}}. \tag{2}$$

For $m \le n/7$, (2) implies that $q_{1/2} = \Omega(2^{\frac{m+n}{2}})$. However, for larger values of m, the bound on $q_{1/2}$ that is offered by (2) deteriorates, and eventually becomes (already for $m > n/4$) worse than the trivial "Birthday" bound $q_{1/2} = \Omega(2^{n/2})$, which is obviously still valid when the adversary gets only partial (truncated) replies from the oracle.

Hall et al. [2] conjectured that $\Omega(2^{\frac{m+n}{2}})$ queries are needed in order to get a non-negligible advantage, in the general case. Surprisingly, it turns out that this was already established 20 years before the conjecture was raised, in a different context. It follows from the bound

$$Adv \le \frac{1}{2}\sqrt{\frac{(2^{n-m} - 1)q(q-1)}{(2^n - 1)(2^n - (q-1))}} \le \frac{1}{2\sqrt{1 - \frac{q-1}{2^n}}} \cdot \frac{q}{2^{\frac{n+m}{2}}}, \tag{3}$$

valid for all $0 \le m < n$, which is a direct consequence of a result of Stam [3, Theorem 2.3] (see also [1]).

2

In this manuscript we show how the method of proof used in [2] can be modified to show the lower bound $q_{1/2} = \Omega(2^{\frac{m+n}{2}})$ for virtually every $m < n$. The result follows from explicit upper bounds on Adv stated in the following two theorems.

Theorem 1. *If $m \leq n/3$ then*

$$Adv \leq 2\sqrt[3]{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + \frac{2\sqrt{2}}{\sqrt{3}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{3/2} + \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2. \tag{4}$$

Theorem 2. *If $n/3 < m \leq n - 4 - \log_2 n$ then*

$$Adv \leq 3\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + 2\left(\frac{q}{2^{\frac{n+m}{2}}}\right) + 5\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + \frac{1}{2}\left(\frac{2q}{2^{\frac{n+m}{2}}}\right)^{\frac{n}{n-m}}. \tag{5}$$

The proofs of Theorem 1 and Theorem 2 are given in Section 3 and Section 4, respectively. The proofs follow the same line, but the proof of Theorem 2 is more elaborate and technical.

2   Notation and preliminaries

For fixed $m < n$ and $q \leq 2^n$ we denote $\Omega := (\{0,1\}^{n-m})^q$. We view Ω as the set of all possible sequences of replies that can be given by the oracle to the adversary's q queries. For $\omega \in \Omega$, let $\Pr_p(\omega)$ and $\Pr_f(\omega)$ be the probabilities that ω is the actual sequence of replies that the oracle gives to the adversary's q queries, in the case the oracle chose a random permutation or a random function, respectively. For every $j \geq 2$ and $\omega \in \Omega$, let

$$\mathrm{col}_j(\omega) := \#\{1 \leq i_1 < i_2 < \ldots < i_j \leq q \mid \omega_{i_1} = \omega_{i_2} = \ldots = \omega_{i_j}\}.$$

Lemma 2.1. *For every $j \geq 2$,*

$$\mathrm{E}_f \, \mathrm{col}_j = \binom{q}{j}\frac{1}{2^{(j-1)(n-m)}},$$

$$\mathrm{Var}_f \, \mathrm{col}_j \leq \binom{j}{2}\binom{q}{j}\frac{1}{2^{(j-1)(n-m)}}\left(1 + \frac{q}{2^{n-m}}\right)^{j-2}.$$

Proof. Note that for every $j \geq 2$,

$$\mathrm{col}_j = \sum_{J \in \binom{[q]}{j}} X_J,$$

where $\binom{[q]}{j} := \{\{i_1, i_2, \ldots, i_j\} \mid 1 \leq i_1 < i_2 < \ldots < i_j \leq q\}$ and $X_{\{i_1,i_2,\ldots,i_j\}}$ is the indicator function of the event $\{\omega_{i_1} = \omega_{i_2} = \ldots = \omega_{i_j}\}$. Since clearly $\mathrm{E}_f \, X_J = (1/2^{n-m})^{|J|-1}$ for every J, we immediately get that

$$\mathrm{E}_f \, \mathrm{col}_j = \mathrm{E}_f \sum_{J \in \binom{[q]}{j}} X_J = \sum_{J \in \binom{[q]}{j}} \mathrm{E}_f \, X_J = \binom{q}{j}\left(\frac{1}{2^{n-m}}\right)^{j-1}.$$

Since X_{J_1} and X_{J_2} are clearly independent whenever J_1 and J_2 are disjoint,

$$\mathrm{Var}_f \, \mathrm{col}_j = \mathrm{Var}_f \sum_{J \in \binom{[q]}{j}} X_J = \sum_{J_1 \in \binom{[q]}{j}} \sum_{\substack{J_2 \in \binom{[q]}{j} \\ J_1 \cap J_2 \neq \emptyset}} \mathrm{E}_f \, X_{J_1} X_{J_2} - \mathrm{E}_f \, X_{J_1} \cdot \mathrm{E}_f \, X_{J_2} =$$

$$= \sum_{J_1 \in \binom{[q]}{j}} \sum_{\substack{J_2 \in \binom{[q]}{j} \\ J_1 \cap J_2 \neq \emptyset}} \left(\frac{1}{2^{n-m}}\right)^{|J_1 \cup J_2| - 1} - \left(\frac{1}{2^{n-m}}\right)^{|J_1| - 1} \left(\frac{1}{2^{n-m}}\right)^{|J_2| - 1} =$$

$$= \binom{q}{j} \frac{1}{2^{(j-1)(n-m)}} \sum_{i=0}^{j-2} \binom{j}{i}\binom{q-j}{i} \left(\frac{1}{2^{i(n-m)}} - \frac{1}{2^{(j-1)(n-m)}}\right) \leq$$

$$\leq \binom{j}{2}\binom{q}{j} \frac{1}{2^{(j-1)(n-m)}} \sum_{i=0}^{j-2} \binom{j-2}{i} q^i \frac{1}{2^{i(n-m)}} = \binom{j}{2}\binom{q}{j} \frac{1}{2^{(j-1)(n-m)}} \left(1 + \frac{q}{2^{n-m}}\right)^{j-2}. \quad \Box$$

The advantage of an algorithm is defined as $\left|\mathrm{Pr}_p(E) - \mathrm{Pr}_f(E)\right|$, where E is the event that the algorithm outputs (say) 1. The maximum of the advantage of an algorithm, over all possible algorithms, is called the adversary's advantage, and is denoted here by Adv. Clearly,

$$Adv \leq \max_{E \subseteq \Omega} |\mathrm{Pr}_p(E) - \mathrm{Pr}_f(E)| = \frac{1}{2} \sum_{\omega \in \Omega} |\mathrm{Pr}_p(\omega) - \mathrm{Pr}_f(\omega)|, \tag{6}$$

with equality, if no computational restrictions are imposed on the adversary. We use the following estimate for Adv, which is slightly better than a similar bound used in [2].

Lemma 2.2. *For every* $S \subseteq \Omega$,

$$Adv \leq \max_{\omega \in S} \left|\frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} - 1\right| + \mathrm{Pr}_f(\bar{S}).$$

Proof. Note that

$$\sum_{\omega \in \bar{S}} |\mathrm{Pr}_p(\omega) - \mathrm{Pr}_f(\omega)| \leq \sum_{\omega \in \bar{S}} (\mathrm{Pr}_p(\omega) + \mathrm{Pr}_f(\omega)) =$$

$$= \mathrm{Pr}_p(\bar{S}) + \mathrm{Pr}_f(\bar{S}) = \mathrm{Pr}_f(S) - \mathrm{Pr}_p(S) + 2\mathrm{Pr}_f(\bar{S}) =$$

$$= \sum_{\omega \in S} (\mathrm{Pr}_f(\omega) - \mathrm{Pr}_p(\omega)) + 2\mathrm{Pr}_f(\bar{S}) \leq \sum_{\omega \in S} |\mathrm{Pr}_p(\omega) - \mathrm{Pr}_f(\omega)| + 2\mathrm{Pr}_f(\bar{S}).$$

Therefore, using (6),

$$Adv \leq \frac{1}{2} \sum_{\omega \in \Omega} |\mathrm{Pr}_p(\omega) - \mathrm{Pr}_f(\omega)| = \frac{1}{2} \sum_{\omega \in S} |\mathrm{Pr}_p(\omega) - \mathrm{Pr}_f(\omega)| + \frac{1}{2} \sum_{\omega \in \bar{S}} |\mathrm{Pr}_p(\omega) - \mathrm{Pr}_f(\omega)| \leq$$

$$\leq \sum_{\omega \in S} |\mathrm{Pr}_p(\omega) - \mathrm{Pr}_f(\omega)| + \mathrm{Pr}_f(\bar{S}) = \sum_{\omega \in S} \mathrm{Pr}_f(\omega) \left|\frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} - 1\right| + \mathrm{Pr}_f(\bar{S}) \leq$$

$$\leq \left(\sum_{\omega \in S} \mathrm{Pr}_f(\omega)\right) \max_{\omega \in S} \left|\frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} - 1\right| + \mathrm{Pr}_f(\bar{S}) \leq \max_{\omega \in S} \left|\frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} - 1\right| + \mathrm{Pr}_f(\bar{S}). \quad \Box$$

In the proofs of Theorem 1 and Theorem 2 we apply Lemma 2.2 to the set

$$S = \left\{ \omega \in \Omega \mid \forall 2 \leq j \leq t : \left| \mathrm{col}_j(\omega) - \binom{q}{j} \frac{1}{2^{(j-1)(n-m)}} \right| \leq \alpha_j,\ \mathrm{col}_{t+1}(\omega) \leq \beta \right\},$$

where $t \geq 2$ is an integer and $\alpha_1, \alpha_2, \ldots, \alpha_{t-1}, \beta$ are positive real numbers, which are chosen apropriately. A particular case of this S, with $t = 2$, $\alpha_1 = c\,q/2^{\frac{n-m+1}{2}}$, $\beta = 0$, was used in [2]. In this work, we get a refined asymptotic approximation for Adv by using the above general choice of S. In the proof of Theorem 1, we also use $t = 2$, but different α_1 (which we simply denote α) and different β.

3   Proof of Theorem 1

The flow of the proof is as follows. As mentioned in Section 2, we let

$$S = \left\{ \omega \in \Omega : \left| \mathrm{col}_2(\omega) - \binom{q}{2} \frac{1}{2^{n-m}} \right| \leq \alpha,\ \mathrm{col}_3(\omega) \leq \beta \right\},$$

where α, β are positive constants to be specifred later. In Subection 3.1 we prove our main technical result, Proposition 3.1, which provides an upper bound for $|\mathrm{Pr}_p/\mathrm{Pr}_f - 1|$ in S. In Subsection 3.2 we first derive, in Lemma 3.5, an upper bound for $\mathrm{Pr}_f(\bar{S})$. Then we combine Lemma 2.2, Proposition 3.1, Lemma 3.5, and choose optimal parameters α, β to obtain Theorem 1.

3.1   Bounding $|\mathrm{Pr}_p/\mathrm{Pr}_f - 1|$ in S

In this subsection, we prove the following proposition.

Proposition 3.1. *Suppose that* $q \leq 2^{n-1}$,

$$\frac{\alpha}{2^m} + \frac{2}{3} \cdot \frac{q^3}{2^{2n}} \leq \frac{1}{2}, \tag{7}$$

$$\binom{q}{2} \frac{1}{2^{n-m}} + \alpha \leq \binom{2^{m-1}}{2}, \tag{8}$$

$$\beta \geq 2 \binom{q}{3} \frac{1}{2^{2(n-m)}}. \tag{9}$$

Then for every $\omega \in S$,

$$\left| \frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} - 1 \right| \leq 2 \frac{\alpha}{2^m} + 2 \binom{q}{2} \frac{1}{2^{n+m}} + 4 \frac{\beta}{2^{2m}}.$$

In the proof of Proposition 3.1, we use the following three lemmas.

Lemma 3.2. *For every* $x \leq 1$,

$$x \leq -\ln(1 - x), \tag{10}$$

for every $0 \leq x \leq 1/2$,

$$-\ln(1 - x) - x \leq 2x^2, \tag{11}$$

and for every $1 \leq x \leq 2$,

$$x - 1 \leq 2 \ln x. \tag{12}$$

Proof. By Taylor expansion, for every $x \leq 1$,

$$-\ln(1-x) = x + \frac{x^2}{2(1-\xi)^2}$$

for some ξ between 0 and x, and (10), (11) follow. To deduce (12), note that if $1 \leq x \leq 2$, then by (10),

$$x - 1 \leq 2\frac{x-1}{x} \leq -2\ln\left(1 - \frac{x-1}{x}\right) = 2\ln x. \qquad \square$$

Lemma 3.3. *Let s, k be positive integers such that $s \leq 2^{k-1}$. Then*

$$0 \leq -\ln \prod_{i=0}^{s-1}\left(1 - \frac{i}{2^k}\right) - \binom{s}{2}\frac{1}{2^k} \leq 4\binom{s}{3}\frac{1}{2^{2k}} + 2\binom{s}{2}\frac{1}{2^{2k}}.$$

Proof. For every integer $0 \leq i \leq 2^{k-1}$, by (10) and (11),

$$0 \leq -\ln\left(1 - \frac{i}{2^k}\right) - \frac{i}{2^k} \leq 2\left(\frac{i}{2^k}\right)^2 = 4\binom{i}{2}\frac{1}{2^{2k}} + 2\frac{i}{2^{2k}}. \qquad (13)$$

The Lemma follows by summing up the inequalities (13) for $0 \leq i \leq s - 1$, and using the identities $\sum_{i=0}^{s-1} i = \binom{s}{2}$, $\sum_{i=0}^{s-1}\binom{i}{2} = \binom{s}{3}$. $\qquad \square$

Lemma 3.4. *Suppose that $q \leq 2^{n-1}$ and let $\omega \in \Omega$ such that $\mathrm{col}_2(\omega) \leq \binom{2^{m-1}}{2}$. Then,*

$$0 \leq -\ln\frac{\Pr_p(\omega)}{\Pr_f(\omega)} - \ln\prod_{i=0}^{q-1}\left(1 - \frac{i}{2^n}\right) - \frac{\mathrm{col}_2(\omega)}{2^m} \leq 4\frac{\mathrm{col}_3(\omega)}{2^{2m}} + 2\frac{\mathrm{col}_2(\omega)}{2^{2m}}.$$

Proof. Suppose that in the q-tuple ω, exactly ℓ distinct vectors in $\{0,1\}^{n-m}$ appear, with multiplicities $d_1, d_2, \ldots, d_\ell$, respectively. It is easy to verify that

$$\Pr_p(\omega) = \frac{\prod_{k=1}^{\ell}\left(\prod_{i=0}^{d_k-1}(2^m - i)\right)}{\prod_{i=0}^{q-1}(2^n - i)},$$

and clearly $\Pr_f(\omega) = (1/2^{n-m})^q$. Therefore, using that $\sum_{k=1}^{\ell} d_k = q$,

$$\frac{\Pr_p(\omega)}{\Pr_f(\omega)} = \prod_{k=1}^{\ell}\left(\prod_{i=0}^{d_k-1}\left(1 - \frac{i}{2^m}\right)\right) \cdot \frac{1}{\prod_{i=0}^{q-1}\left(1 - \frac{i}{2^n}\right)},$$

hence

$$\ln\frac{\Pr_p(\omega)}{\Pr_f(\omega)} + \ln\prod_{i=0}^{q-1}\left(1 - \frac{i}{2^n}\right) = \sum_{k=1}^{\ell}\ln\prod_{i=0}^{d_k-1}\left(1 - \frac{i}{2^m}\right). \qquad (14)$$

For every $1 \leq k \leq \ell$, note that $d_k \leq 2^{m-1}$, since $\sum_{k=1}^{\ell}\binom{d_k}{2} = \mathrm{col}_2(\omega) \leq \binom{2^{m-1}}{2}$, hence by Lemma 3.3,

$$0 \leq -\ln\prod_{i=0}^{d_k-1}\left(1 - \frac{i}{2^m}\right) - \binom{d_k}{2}\frac{1}{2^m} \leq 4\binom{d_k}{3}\frac{1}{2^{2m}} + 2\binom{d_k}{2}\frac{1}{2^{2m}}.$$

Summing up on $1 \leq k \leq \ell$, we get that

$$0 \leq -\sum_{k=1}^{\ell}\ln\prod_{i=0}^{d_k-1}\left(1 - \frac{i}{2^m}\right) - \frac{1}{2^m}\sum_{k=1}^{\ell}\binom{d_k}{2} \leq \frac{4}{2^{2m}}\sum_{k=1}^{\ell}\binom{d_k}{3} + \frac{2}{2^{2m}}\sum_{k=1}^{\ell}\binom{d_k}{2},$$

and the lemma follows by (14) since $\sum_{k=1}^{\ell}\binom{d_k}{2} = \mathrm{col}_2(\omega)$ and $\sum_{k=1}^{\ell}\binom{d_k}{3} = \mathrm{col}_3(\omega)$. $\qquad \square$

6

We are now ready to prove Proposition 3.1.

Proof of Proposition 3.1. For every $\omega \in S$, by (8) and the definition of S,

$$\text{col}_2(\omega) \le \binom{q}{2}\frac{1}{2^{n-m}} + \alpha \le \binom{2^{m-1}}{2},$$

hence, by Lemma 3.4,

$$0 \le -\ln\frac{\text{Pr}_p(\omega)}{\text{Pr}_f(\omega)} - \ln\prod_{i=0}^{q-1}\left(1 - \frac{i}{2^n}\right) - \frac{\text{col}_2(\omega)}{2^m} \le 4\frac{\text{col}_3(\omega)}{2^{2m}} + 2\frac{\text{col}_2(\omega)}{2^{2m}}.$$

In addition, by Lemma 3.3 (for $s = q$ and $k = n$), since $q \le 2^{n-1}$,

$$0 \le -\ln\prod_{i=0}^{q-1}\left(1 - \frac{i}{2^n}\right) - \frac{1}{2^n}\binom{q}{2} \le 4\binom{q}{3}\frac{1}{2^{2n}} + 2\binom{q}{2}\frac{1}{2^{2n}}.$$

Therefore,

$$-\ln\frac{\text{Pr}_p(\omega)}{\text{Pr}_f(\omega)} - \frac{1}{2^m}\left(\text{col}_2(\omega) - \binom{q}{2}\frac{1}{2^{n-m}}\right) \le 4\frac{\text{col}_3(\omega)}{2^{2m}} + 2\frac{\text{col}_2(\omega)}{2^{2m}}, \tag{15}$$

$$\ln\frac{\text{Pr}_p(\omega)}{\text{Pr}_f(\omega)} + \frac{1}{2^m}\left(\text{col}_2(\omega) - \binom{q}{2}\frac{1}{2^{n-m}}\right) \le 4\binom{q}{3}\frac{1}{2^{2n}} + 2\binom{q}{2}\frac{1}{2^{2n}}. \tag{16}$$

By (10), (15) and the definition of S,

$$1 - \frac{\text{Pr}_p(\omega)}{\text{Pr}_f(\omega)} \le -\ln\frac{\text{Pr}_p(\omega)}{\text{Pr}_f(\omega)} \le \frac{1}{2^m}\left(\text{col}_2(\omega) - \binom{q}{2}\frac{1}{2^{n-m}}\right) + 2\frac{\text{col}_2(\omega)}{2^{2m}} + 4\frac{\text{col}_3(\omega)}{2^{2m}} \le$$

$$\le \frac{\alpha}{2^m} + \frac{2}{2^{2m}}\left(\binom{q}{2}\frac{1}{2^{n-m}} + \alpha\right) + 4\frac{\beta}{2^{2m}} = \left(1 + \frac{2}{2^m}\right)\frac{\alpha}{2^m} + 2\binom{q}{2}\frac{1}{2^{n+m}} + 4\frac{\beta}{2^{2m}} \le$$

$$\le 2\frac{\alpha}{2^m} + 2\binom{q}{2}\frac{1}{2^{n+m}} + 4\frac{\beta}{2^{2m}}. \tag{17}$$

By (16) and the definition of S,

$$\ln\frac{\text{Pr}_p(\omega)}{\text{Pr}_f(\omega)} \le -\frac{1}{2^m}\left(\text{col}_2(\omega) - \binom{q}{2}\frac{1}{2^{n-m}}\right) + 4\binom{q}{3}\frac{1}{2^{2n}} + 2\binom{q}{2}\frac{1}{2^{2n}} \le$$

$$\le \frac{\alpha}{2^m} + 4\binom{q}{3}\frac{1}{2^{2n}} + 2\binom{q}{2}\frac{1}{2^{2n}}. \tag{18}$$

In particular, using (7),

$$\frac{\text{Pr}_p(\omega)}{\text{Pr}_f(\omega)} \le \exp\left(\frac{\alpha}{2^m} + 4\binom{q}{3}\frac{1}{2^{2n}} + 2\binom{q}{2}\frac{1}{2^{2n}}\right) < \exp\left(\frac{\alpha}{2^m} + \frac{2}{3}\cdot\frac{q^3}{2^{2n}}\right) < 2,$$

hence, if $\text{Pr}_p(\omega)/\text{Pr}_f(\omega) \ge 1$ then by (12), (18) and (9),

$$\frac{\text{Pr}_p(\omega)}{\text{Pr}_f(\omega)} - 1 \le 2\left(\frac{\alpha}{2^m} + 4\binom{q}{3}\frac{1}{2^{2n}} + 2\binom{q}{2}\frac{1}{2^{2n}}\right) =$$

$$= 2\frac{\alpha}{2^m} + \frac{4}{2^{n-m}}\binom{q}{2}\frac{1}{2^{n+m}} + 4\cdot2\binom{q}{3}\frac{1}{2^{2(n-m)}}\cdot\frac{1}{2^{2m}} \le 2\frac{\alpha}{2^m} + 2\binom{q}{2}\frac{1}{2^{n+m}} + 4\frac{\beta}{2^{2m}}. \tag{19}$$

The proposition now follows from (17) and (19). $\qquad\square$

3.2 Derivation of Theorem 1

We start by bounding $\Pr_f(\bar{S})$ from above.

Lemma 3.5.

$$\Pr_f(\bar{S}) \leq \binom{q}{2} \frac{1}{2^{n-m}} \cdot \frac{1}{\alpha^2} + \binom{q}{3} \frac{1}{2^{2(n-m)}} \cdot \frac{1}{\beta} \leq \frac{1}{2} \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^2 \left(\frac{2^m}{\alpha} \right)^2 + \frac{1}{6 \cdot 2^{\frac{n-3m}{2}}} \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^3 \frac{2^{2m}}{\beta}.$$

Proof. By Chebyshev inequality,

$$\Pr_f \left(\left\{ \omega \in \Omega : \left| \operatorname{col}_2(\omega) - \binom{q}{2} \frac{1}{2^{n-m}} \right| > \alpha \right\} \right) \leq \frac{\operatorname{Var}_f \operatorname{col}_2}{\alpha^2},$$

and by Markov inequality,

$$\Pr_f \left(\{ \omega \in \Omega : \operatorname{col}_3(\omega) > \beta \} \right) \leq \frac{\operatorname{E}_f \operatorname{col}_3}{\beta}.$$

Using the union bound, we conclude that

$$\Pr_f(\bar{S}) \leq \frac{\operatorname{Var}_f \operatorname{col}_2}{\alpha^2} + \frac{\operatorname{E}_f \operatorname{col}_3}{\beta},$$

and the claim follows by Lemma 2.1. $\qquad\square$

We now combine Lemma 2.2, Lemma 3.5 and Proposition 3.1.

Lemma 3.6. *Suppose that $q \leq 2^{n-1}$,*

$$\frac{\alpha}{2^m} + \frac{2}{3} \cdot \frac{1}{2^{\frac{n-3m}{2}}} \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^3 \leq \frac{1}{2},$$

$$\left(\frac{q}{2^{\frac{n+m}{2}}} \right)^2 + \frac{2\alpha}{2^{2m}} \leq \frac{1}{2} \left(\frac{1}{2} - \frac{1}{2^m} \right),$$

$$\beta \geq 2 \binom{q}{3} \frac{1}{2^{2(n-m)}}.$$

Then

$$Adv \leq \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^2 + \left(2 \frac{\alpha}{2^m} + \frac{1}{2} \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^2 \left(\frac{2^m}{\alpha} \right)^2 \right) +$$

$$+ \left(4 \frac{\beta}{2^{2m}} + \frac{1}{6 \cdot 2^{\frac{n-3m}{2}}} \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^3 \frac{2^{2m}}{\beta} \right). \quad (20)$$

Proof. All the conditions of Proposition 3.1 are clearly satisfied. Therefore,

$$\max_{\omega \in S} \left| \frac{\Pr_p(\omega)}{\Pr_f(\omega)} - 1 \right| \leq 2 \frac{\alpha}{2^m} + 2 \binom{q}{2} \frac{1}{2^{n+m}} + 4 \frac{\beta}{2^{2m}} \leq 2 \frac{\alpha}{2^m} + \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^2 + 4 \frac{\beta}{2^{2m}},$$

and the claim follows by Lemma 2.2 and Lemma 3.5. $\qquad\square$

Theorem 1 now follows by taking α, β to minimize the right hand side of (20).

8

Proof of Theorem 1. If $q \geq \frac{1}{4}2^{\frac{m+n}{2}}$ then (4) holds since surely $Adv \leq 1$ and

$$1 = 2\sqrt[3]{2}\left(\frac{1}{4}\right)^{2/3} \leq 2\sqrt[3]{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} < 2\sqrt[3]{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + \frac{2\sqrt{2}}{\sqrt{3}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{3/2} + \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 .$$

If $m \leq 5$ and $q < \frac{1}{4}2^{\frac{m+n}{2}}$, then (4) also holds, since by the "Birthday" bound (1), which is obviously still valid when the adversary gets only truncated replies from the oracle,

$$Adv < \frac{q^2}{2^{n+1}} = 2^{m-5}\left(\frac{4q}{2^{\frac{m+n}{2}}}\right)^2 < \left(\frac{4q}{2^{\frac{m+n}{2}}}\right)^{2/3} = 2\sqrt[3]{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} <$$

$$< 2\sqrt[3]{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + \frac{2\sqrt{2}}{\sqrt{3}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{3/2} + \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 .$$

Finally, if $6 \leq m \leq \frac{n}{3}$ and $q < \frac{1}{4}2^{\frac{m+n}{2}}$, then it is straightforward to verify that all the conditions of Lemma 3.6 are satisfied if we choose

$$\alpha := \frac{2^m}{\sqrt[3]{4}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} , \quad \beta := \frac{2^{2m}}{2\sqrt{6}\cdot 2^{\frac{n-3m}{4}}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{3/2} .$$

Then, by Lemma 3.6,

$$Adv \leq \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + \left(2\frac{\alpha}{2^m} + \frac{1}{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2\left(\frac{2^m}{\alpha}\right)^2\right) + \left(4\frac{\beta}{2^{2m}} + \frac{1}{6\cdot 2^{\frac{n-3m}{2}}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^3\frac{2^{2m}}{\beta}\right) =$$

$$= \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + 2\sqrt[3]{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + \frac{2\sqrt{2}}{\sqrt{3}\cdot 2^{\frac{n-3m}{4}}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{3/2} \leq$$

$$\leq 2\sqrt[3]{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + \frac{2\sqrt{2}}{\sqrt{3}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{3/2} + \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 . \square$$

4 Proof of Theorem 2

The proof of Theorem 2 is more elaborate and technical than the proof of Theorem 1, but goes along a similar path. It uses statements that are analogous to those used in the proof of Theorem 1. As mentioned in Section 2, we let

$$S = \left\{\omega \in \Omega \mid \forall 2 \leq j \leq t : \left|\mathrm{col}_j(\omega) - \binom{q}{j}\frac{1}{2^{(j-1)(n-m)}}\right| \leq \alpha_j, \ \mathrm{col}_{t+1}(\omega) \leq \beta\right\},$$

where $t \geq 2$ is an integer and $\alpha_1, \alpha_2, \ldots, \alpha_{t-1}, \beta$ are positive real numbers to be specifired later.

4.1 Bounding $|\mathrm{Pr}_p/\mathrm{Pr}_f - 1|$ in S

The following proposition is analogous to Proposition 3.1.

Proposition 4.1. *Suppose that $m \leq n - 2$, $q \leq 2^{n-1}$, $2 \leq t \leq 2^{(m-1)/2} + 1$,*

$$\binom{q}{2}\frac{1}{2^{n-m}} + \alpha_1 \leq \binom{2^{m-1}}{2}, \tag{21}$$

$$4\left(\frac{q}{2^n}\right)^2 + \frac{1}{2t(t+1)2^{\frac{n-m}{2}(t-\frac{n+m}{n+m})}}\left(\frac{2(q+t-1)}{2^{\frac{n+m}{2}}}\right)^{t+1} + \sum_{j=1}^{t-1}\frac{(j-1)!}{2^{jm}}\alpha_j \leq \frac{1}{2}, \tag{22}$$

$$\beta \geq 2\binom{q}{t+1}\frac{1}{2^{t(n-m)}}. \tag{23}$$

Then for every $\omega \in S$,

$$\left| \frac{\Pr_p(\omega)}{\Pr_f(\omega)} - 1 \right| \leq 4 \left(1 + \left(\frac{2(t-1)}{2^m} + \frac{2q}{2^n} \right)^{t-2} \right) \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^2 +$$

$$+ 2 \left(1 + \frac{4}{2^m} \right) \frac{\alpha_1}{2^m} + \left(\sum_{j=2}^{t-1} (1 + 2^j) \frac{(j-1)!}{2^{jm}} \alpha_j \right) + \frac{2^t(t-1)!}{2^{tm}} \beta.$$

In the proof of the proposition we use several lemmas.

Lemma 4.2. *For every integers $i \geq 0$, $j \geq 1$,*

$$0 \leq \frac{i^j}{j!} - \binom{i}{j} \leq \sum_{r=1}^{j-1} \binom{j-1}{r-1} \binom{i}{r}.$$

Proof. Note that
$$i(i-1)\ldots(i-(j-1)) \leq i^j \leq (i+j-1)\ldots(i+1)i,$$

hence

$$\binom{i}{j} \leq \frac{i^j}{j!} \leq \binom{i+j-1}{j} = \sum_{r=1}^{j} \binom{j-1}{j-r} \binom{i}{r},$$

and the claim follows. □

With this we prove the following lemma.

Lemma 4.3. *Let i, k, t be integers such that $k \geq 1$, $0 \leq i \leq 2^{k-1}$, $t \geq 2$. Then,*

$$0 \leq -\ln\left(1 - \frac{i}{2^k}\right) - \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \binom{i}{j} \leq$$

$$\leq \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \sum_{r=1}^{j-1} \binom{j-1}{r-1} \binom{i}{r} + \frac{2^t(t-1)!}{2^{tk}} \sum_{r=1}^{t} \binom{t-1}{r-1} \binom{i}{r}. \quad (24)$$

Proof. By Taylor expansion, for every $x > 0$

$$-\ln(1-x) = \sum_{j=1}^{t-1} \frac{x^j}{j} + \frac{x^t}{t(1-\xi)^t}$$

for some $0 < \xi < x$. It follows that for every $0 \leq x \leq \frac{1}{2}$,

$$0 \leq -\ln(1-x) - \sum_{j=1}^{t-1} \frac{x^j}{j} \leq \frac{2^t}{t} x^t.$$

In particular,

$$0 \leq -\ln\left(1 - \frac{i}{2^k}\right) - \sum_{j=1}^{t-1} \frac{1}{j} \left(\frac{i}{2^k}\right)^j \leq \frac{2^t}{t} \left(\frac{i}{2^k}\right)^t. \quad (25)$$

10

and by Lemma 4.2,

$$0 \le \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \left(\frac{i^j}{j!} - \binom{i}{j} \right) \le \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \sum_{r=1}^{j-1} \binom{j-1}{r-1} \binom{i}{r},$$

hence

$$0 \le -\ln\left(1 - \frac{i}{2^k}\right) - \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \binom{i}{j} \le \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \sum_{r=1}^{j-1} \binom{j-1}{r-1} \binom{i}{r} + \frac{2^t}{t} \left(\frac{i}{2^k} \right)^t,$$

and the lemma follows since by Lemma 4.2,

$$\frac{2^t}{t} \left(\frac{i}{2^k} \right)^t = \frac{2^t(t-1)!}{2^{tk}} \cdot \frac{i^t}{t!} \le \frac{2^t(t-1)!}{2^{tk}} \sum_{r=1}^{t} \binom{t-1}{r-1} \binom{i}{r}. \qquad \square$$

This leads to the following lemma, which is a generalization of Lemma 3.3.

Lemma 4.4. *Let s, k, t be positive integers such that $s \le 2^{k-1}$, $2 \le t \le 2^{k/2} + 2$. Then,*

$$0 \le -\ln \prod_{i=0}^{s-1} \left(1 - \frac{i}{2^k}\right) - \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \binom{s}{j+1} \le 8\binom{s}{2} \frac{1}{2^{2k}} + \frac{2^t(t-1)!}{2^{tk}} \sum_{j=1}^{t} \binom{t-1}{j-1} \binom{s}{j+1}.$$

Proof. By summing up (24) for $0 \le i \le s-1$ and using the identity $\sum_{i=0}^{s-1} \binom{i}{j} = \binom{s}{j+1}$ we get that

$$0 \le -\ln \prod_{i=0}^{s-1} \left(1 - \frac{i}{2^k}\right) - \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \binom{s}{j+1} \le$$

$$\le \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \sum_{r=1}^{j-1} \binom{j-1}{r-1} \binom{s}{r+1} + \frac{2^t(t-1)!}{2^{tk}} \sum_{r=1}^{t} \binom{t-1}{r-1} \binom{s}{r+1}.$$

For every r, j such that $2 \le r + 1 \le j < t - 1$,

$$\frac{j!}{2^{(j+1)k}} \binom{j}{r-1} \bigg/ \frac{(j-1)!}{2^{jk}} \binom{j-1}{r-1} = \frac{j}{2^k} \cdot \frac{j}{j-(r-1)} \le \frac{(t-2)^2}{2^k \cdot 2} \le \frac{1}{2},$$

hence, for every $1 \le r \le t - 2$,

$$\sum_{j=r+1}^{t-1} \frac{(j-1)!}{2^{jk}} \binom{j-1}{r-1} < 2 \cdot \frac{r! \cdot r}{2^{rk}}.$$

Therefore,

$$\sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jk}} \sum_{r=1}^{j-1} \binom{j-1}{r-1} \binom{s}{r+1} = \sum_{r=1}^{t-2} \left(\sum_{j=r+1}^{t-1} \frac{(j-1)!}{2^{jk}} \binom{j-1}{r-1} \right) \binom{s}{r+1} <$$

$$< \sum_{r=1}^{t-2} 2 \frac{r! \cdot r}{2^{(r+1)k}} \binom{s}{r+1} \le \sum_{r=1}^{t-2} 2 \frac{s(s-1)^r}{2^{(r+1)k}} = 4\binom{s}{2} \frac{1}{2^{2k}} \sum_{r=1}^{t-2} \left(\frac{s-1}{2^k} \right)^{r-1} < 8\binom{s}{2} \frac{1}{2^{2k}},$$

and the lemma follows. $\qquad \square$

11

This leads to the following lemma, which is a generalization of Lemma 3.4.

Lemma 4.5. *If $q \leq 2^{n-1}$, $2 \leq t \leq 2^{m/2} + 2$, then for every $\omega \in \Omega$ for which $\mathrm{col}_2(\omega) \leq \binom{2^{m-1}}{2}$,*

$$0 \leq -\ln \frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} - \ln \prod_{i=0}^{q-1} \left(1 - \frac{i}{2^n}\right) - \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \mathrm{col}_{j+1}(\omega) \leq$$

$$\leq 8 \frac{\mathrm{col}_2(\omega)}{2^{2m}} + \frac{2^t(t-1)!}{2^{tm}} \sum_{j=1}^{t} \binom{t-1}{j-1} \mathrm{col}_{j+1}(\omega).$$

Proof. Suppose that in the q-tuple ω exactly t distinct vectors in $\{0,1\}^{n-m}$ appear, with multiplicities $d_1, d_2, \ldots, d_l$ respectively. For every $1 \leq k \leq l$, note that $d_k \leq 2^{m-1}$, since $\sum_{k=1}^{l} \binom{d_k}{2} = \mathrm{col}_2(\omega) \leq \binom{2^{m-1}}{2}$, hence by Lemma 4.4,

$$0 \leq -\ln \prod_{i=0}^{d_k-1} \left(1 - \frac{i}{2^m}\right) - \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \binom{d_k}{j+1} \leq 8 \binom{d_k}{2} \frac{1}{2^{2m}} + \frac{2^t(t-1)!}{2^{tm}} \sum_{j=1}^{t} \binom{t-1}{j-1} \binom{d_k}{j+1}.$$

Summing up on $1 \leq k \leq l$, we get

$$0 \leq -\sum_{k=1}^{l} \ln \prod_{i=0}^{d_k-1} \left(1 - \frac{i}{2^m}\right) - \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \sum_{k=1}^{l} \binom{d_k}{j+1} \leq$$

$$\leq \frac{8}{2^{2m}} \sum_{k=1}^{l} \binom{d_k}{2} + \frac{2^t(t-1)!}{2^{tm}} \sum_{j=1}^{t} \binom{t-1}{j-1} \sum_{k=1}^{l} \binom{d_k}{j+1},$$

and the claim follows by (14). $\qquad\square$

Finally, we also need the following technical lemma.

Lemma 4.6. *For every integers n, m, q and t*

$$\frac{2^t(t-1)!}{2^{tm}} \sum_{j=1}^{t-1} \binom{t-1}{j-1} \binom{q}{j+1} \frac{1}{2^{j(n-m)}} \leq 4 \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 \left(\frac{2(t-1)}{2^m} + \frac{2q}{2^n}\right)^{t-2}.$$

Proof. Note that

$$(t-1)! \sum_{j=1}^{t-1} \binom{t-1}{j-1} \binom{q}{j+1} \frac{1}{2^{j(n-m)}} \leq (t-1)! \sum_{j=1}^{t-1} (t-1) \binom{t-2}{j-1} \frac{q^{j+1}}{(j+1)!} \frac{1}{2^{j(n-m)}} =$$

$$= \frac{q^2}{2^{n-m}} \sum_{j=1}^{t-1} \binom{t-2}{j-1} \frac{(t-1)(t-1)!}{(j+1)!} \left(\frac{q}{2^{n-m}}\right)^{j-1} \leq$$

$$\leq \frac{q^2}{2^{n-m}} \sum_{j=1}^{t-1} \binom{t-2}{j-1} (t-1)^{t-j-1} \left(\frac{q}{2^{n-m}}\right)^{j-1} = \frac{q^2}{2^{n-m}} \left((t-1) + \frac{q}{2^{n-m}}\right)^{t-2},$$

hence

$$\frac{2^t(t-1)!}{2^{tm}} \sum_{j=1}^{t-1} \binom{t-1}{j-1} \binom{q}{j+1} \frac{1}{2^{j(n-m)}} \leq \frac{2^t}{2^{tm}} \cdot \frac{q^2}{2^{n-m}} \left((t-1) + \frac{q}{2^{n-m}}\right)^{t-2} =$$

$$= 4 \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 \left(\frac{2(t-1)}{2^m} + \frac{2q}{2^n}\right)^{t-2}. \qquad\square$$

Proof of Proposition 4.1. For every $\omega \in S$, by (21),

$$\mathrm{col}_2(\omega) \leq \binom{q}{2} \frac{1}{2^{n-m}} + \alpha_1 \leq \binom{2^{m-1}}{2},$$

hence by Lemma 4.5 and Lemma 4.4,

$$-\ln \frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} \leq 8 \frac{\mathrm{col}_2(\omega)}{2^{2m}} + \frac{2^t(t-1)!}{2^{tm}} \sum_{j=1}^{t} \binom{t-1}{j-1} \mathrm{col}_{j+1}(\omega) +$$

$$+ \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \left(\mathrm{col}_{j+1}(\omega) - \binom{q}{j+1} \frac{1}{2^{j(n-m)}} \right), \quad (26)$$

$$\ln \frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} \leq 8 \binom{q}{2} \frac{1}{2^{2n}} + \frac{2^t(t-1)!}{2^{tn}} \sum_{j=1}^{t} \binom{t-1}{j-1} \binom{q}{j+1} -$$

$$- \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \left(\mathrm{col}_{j+1}(\omega) - \binom{q}{j+1} \frac{1}{2^{j(n-m)}} \right). \quad (27)$$

By (10), (26) and the definition of S,

$$1 - \frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} \leq -\ln \frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} \leq$$

$$\leq \frac{8}{2^{2m}} \left(\binom{q}{2} \frac{1}{2^{n-m}} + \alpha_1 \right) +$$

$$+ \frac{2^t(t-1)!}{2^{tm}} \left(\sum_{j=1}^{t-1} \binom{t-1}{j-1} \left(\binom{q}{j+1} \frac{1}{2^{j(n-m)}} + \alpha_j \right) + \beta \right) + \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \alpha_j =$$

$$= 8 \binom{q}{2} \frac{1}{2^{n+m}} + \frac{2^t(t-1)!}{2^{tm}} \sum_{j=1}^{t-1} \binom{t-1}{j-1} \binom{q}{j+1} \frac{1}{2^{j(n-m)}} +$$

$$+ 8 \frac{\alpha_1}{2^{2m}} + \left(\sum_{j=1}^{t-1} \left(\frac{(j-1)!}{2^{jm}} + \frac{2^t(t-1)!}{2^{tm}} \binom{t-1}{j-1} \right) \alpha_j \right) + \frac{2^t(t-1)!}{2^{tm}} \beta \leq$$

$$\leq 4 \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^2 + 4 \left(\frac{q}{2^{\frac{n+m}{2}}} \right)^2 \left(\frac{2(t-1)}{2^m} + \frac{2q}{2^n} \right)^{t-2} +$$

$$+ \left(2 + \frac{8}{2^m} \right) \frac{\alpha_1}{2^m} + \left(\sum_{j=2}^{t-1} (1+2^j) \frac{(j-1)!}{2^{jm}} \alpha_j \right) + \frac{2^t(t-1)!}{2^{tm}} \beta, \quad (28)$$

where on the last step we used Lemma 4.6 and the fact that for every $1 \leq j \leq t-1$, since $(t-1)^2 \leq 2^{m-1}$,

$$\frac{2^t(t-1)!}{2^{tm}} \binom{t-1}{j-1} \leq \frac{2^t(t-1)!}{2^{tm}} \cdot \frac{(t-1)!}{(j-1)!} = \frac{2^t(j-1)!}{2^{tm}} \left(\frac{(t-1)!}{(j-1)!} \right)^2 \leq$$

$$\leq \frac{2^t(j-1)!}{2^{tm}} (t-1)^{2(t-j)} = 2^j \frac{(j-1)!}{2^{jm}} \left(\frac{(t-1)^2}{2^{m-1}} \right)^{t-j} \leq 2^j \frac{(j-1)!}{2^{jm}}.$$

On the other hand, by (27) and the definition of S,

$$\ln \frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} \leq 8 \binom{q}{2} \frac{1}{2^{2n}} + \frac{2^t(t-1)!}{2^{tn}} \sum_{j=1}^{t} \binom{t-1}{j-1} \binom{q}{j+1} + \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \alpha_j. \quad (29)$$

13

In particular, by (22)

$$\ln \frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} \le 8 \binom{q}{2} \frac{1}{2^{2n}} + \frac{2^t(t-1)!}{2^{tn}} \binom{q+t-1}{t+1} + \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \alpha_j \le$$

$$\le 4\left(\frac{q}{2^n}\right)^2 + \frac{1}{2t(t+1)2^{\frac{n-m}{2}(t-\frac{n+m}{n+m})}} \left(\frac{2(q+t-1)}{2^{\frac{n+m}{2}}}\right)^{t+1} + \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}} \alpha_j \le \frac{1}{2},$$

hence $\mathrm{Pr}_p(\omega)/\mathrm{Pr}_f(\omega) \le \sqrt{e} < 2$. Therefore, if $\mathrm{Pr}_p(\omega)/\mathrm{Pr}_f(\omega) \ge 1$ then by (12), (29) and (23),

$$\frac{\mathrm{Pr}_p(\omega)}{\mathrm{Pr}_f(\omega)} - 1 \le 2\left(8\binom{q}{2}\frac{1}{2^{2n}} + \frac{2^t(t-1)!}{2^{tn}} \sum_{j=1}^t \binom{t-1}{j-1}\binom{q}{j+1} + \sum_{j=1}^{t-1} \frac{(j-1)!}{2^{jm}}\alpha_j\right) \le$$

$$\le 16\frac{q^2}{2^{2n}} + 2\frac{2^t(t-1)!}{2^{tn}} \sum_{j=1}^{t-1}\binom{t-1}{j-1}\binom{q}{j+1}2^{(t-j-1)(n-m)}+$$

$$+ 2\frac{2^t(t-1)!}{2^{tn}}2^{t(n-m)}\beta + 2\sum_{j=1}^{t-1}\frac{(j-1)!}{2^{jm}}\alpha_j =$$

$$= \frac{4}{2^{n-m-2}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + \frac{2}{2^{n-m}}\left(\frac{2^t(t-1)!}{2^{tm}}\sum_{j=1}^{t-1}\binom{t-1}{j-1}\binom{q}{j+1}\frac{1}{2^{j(n-m)}}\right)+$$

$$+ \sum_{j=1}^{t-1}2\frac{(j-1)!}{2^{jm}}\alpha_j + \frac{2^t(t-1)!}{2^{tm}}\beta,$$

and we are done by Lemma 4.6. $\qquad\square$

4.2 Derivation of Theorem 2

The following lemma is a generalization of Lemma 3.5.

Lemma 4.7.

$$\mathrm{Pr}_f(\bar{S}) \le \sum_{j=1}^{t-1}\binom{j+1}{2}\binom{q}{j+1}\frac{1}{2^{j(n-m)}}\left(1+\frac{q}{2^{n-m}}\right)^{j-1}\frac{1}{\alpha_j^2} + \binom{q}{t+1}\frac{1}{2^{t(n-m)}\beta}.$$

Proof. By Chebyshev inequality, for every $1 \le j \le t-1$,

$$\mathrm{Pr}_f\left(\left\{\omega \in \Omega : \left|\mathrm{col}_{j+1}(\omega) - \binom{q}{j+1}\frac{1}{2^{j(n-m)}}\right| > \alpha_j\right\}\right) \le \frac{\mathrm{Var}_f \, \mathrm{col}_{j+1}}{\alpha_j^2},$$

and by Markov inequality,

$$\mathrm{Pr}_f\left(\{\omega \in \Omega : \mathrm{col}_{t+1}(\omega) > \beta\}\right) \le \frac{\mathrm{E}_f \, \mathrm{col}_{t+1}}{\beta}.$$

Using the union bound, we conclude that

$$\mathrm{Pr}_f(\bar{S}) \le \left(\sum_{j=1}^{t-1}\frac{\mathrm{Var}_f \, \mathrm{col}_{j+1}}{\alpha_j^2}\right) + \frac{\mathrm{E}_f \, \mathrm{col}_{t+1}}{\beta},$$

and the claim follows by Lemma 2.1. $\qquad\square$

14

Combining Lemma 2.2, Lemma 4.7 and Proposition 4.1, we get the following generalization of Lemma 3.6.

Lemma 4.8. *Suppose that $m \leq n - 2$, $q \leq 2^{n-1}$, $2 \leq t \leq 2^{(m-1)/2} + 1$,*

$$\binom{q}{2}\frac{1}{2^{n-m}} + \alpha_1 \leq \binom{2^{m-1}}{2},$$

$$4\left(\frac{q}{2^n}\right)^2 + \frac{1}{2t(t+1)2^{\frac{n-m}{2}(t-\frac{n+m}{n+m})}}\left(\frac{2(q+t-1)}{2^{\frac{n+m}{2}}}\right)^{t+1} + \sum_{j=1}^{t-1}\frac{(j-1)!}{2^{jm}}\alpha_j \leq \frac{1}{2},$$

$$\beta \geq 2\binom{q}{t+1}\frac{1}{2^{t(n-m)}}.$$

Then

$$Adv \leq 4\left(1 + \left(\frac{2(t-1)}{2^m} + \frac{2q}{2^n}\right)^{t-2}\right)\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + \left(2\left(1 + \frac{4}{2^m}\right)\frac{\alpha_1}{2^m} + \frac{1}{2}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2\left(\frac{2^m}{\alpha_1}\right)^2\right) +$$

$$+\sum_{j=2}^{t-1}\left((1+2^j)\frac{(j-1)!\alpha_j}{2^{jm}} + \frac{(j-1)!}{2\cdot 2^{(j-1)(n-m)}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{j+1}\left(\frac{1}{2^{\frac{3m-n}{2}}} + \frac{q}{2^{\frac{n+m}{2}}}\right)^{j-1}\left(\frac{2^{jm}}{(j-1)!\alpha_j}\right)^2\right) +$$

$$+\left(\frac{2^t(t-1)!}{2^{tm}}\beta + \frac{q^{t+1}}{(t+1)!}\cdot\frac{1}{2^{t(n-m)}\beta}\right). \quad (30)$$

Taking $\alpha_1, \alpha_2, \ldots, \alpha_{t-1}, \beta$ to minimize the right hand side of (30) we get the following.

Lemma 4.9. *Assume that $m \leq n - 2$, $q \leq 2^{n-1}$, $2 \leq t \leq 2^{(m-1)/2} + 1$,*

$$\frac{1}{2^{\frac{n-m}{2}(t-\frac{n+m}{n-m})}}\left(\frac{2q}{2^{\frac{n+m}{2}}}\right)^{t+1} \leq \frac{t(t+1)}{4},$$

$$\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + \frac{1}{\sqrt[3]{1+\frac{4}{2^m}}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3}\frac{1}{2^m} \leq \frac{1}{2}\left(\frac{1}{2} - \frac{1}{2^m}\right),$$

$$\frac{4(t-2)}{2^{n-m}}\left(\frac{1}{2^{\frac{3m-n}{2}}} + \frac{q}{2^{\frac{n+m}{2}}}\right)\frac{q}{2^{\frac{n+m}{2}}} \leq \frac{1}{8},$$

and

$$4\left(\frac{q}{2^n}\right)^2 + \frac{1}{2t(t+1)2^{\frac{n-m}{2}(t-\frac{n+m}{n+m})}}\left(\frac{2(q+t-1)}{2^{\frac{n+m}{2}}}\right)^{t+1} +$$

$$+ \frac{1}{2\sqrt[3]{1+\frac{4}{2^m}}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + \left(\frac{2}{5\cdot 2^{n-m}}\left(\frac{1}{2^{\frac{3m-n}{2}}} + \frac{q}{2^{\frac{n+m}{2}}}\right)\right)^{\frac{1}{3}}\frac{q}{2^{\frac{n+m}{2}}} \leq \frac{1}{2}.$$

Then

$$Adv \leq 4\left(1 + \left(\frac{2(t-1)}{2^m} + \frac{2q}{2^n}\right)^{t-2}\right)\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + 2\sqrt[3]{2}\left(1 + \frac{4}{2^m}\right)^{2/3}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} +$$

$$+ 2\sqrt[3]{100}\left(\frac{1}{2^{n-m}}\left(\frac{1}{2^{\frac{3m-n}{2}}} + \frac{q}{2^{\frac{n+m}{2}}}\right)\right)^{\frac{1}{3}}\frac{q}{2^{\frac{n+m}{2}}} + \frac{\sqrt{2}}{\sqrt{t(t+1)}}\cdot\frac{1}{2^{\frac{n-m}{4}(t-\frac{n+m}{n-m})}}\left(\frac{2q}{2^{\frac{n+m}{2}}}\right)^{\frac{t+1}{2}}. \quad (31)$$

Proof. Take

$$\alpha_1 := \frac{2^m}{\sqrt[3]{4\left(1+\frac{4}{2^m}\right)}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3},$$

$$\alpha_j := \frac{2^{jm}}{\sqrt[3]{2\left(1+2^j\right)(j-1)!^2}}\left(\frac{1}{2^{(j-1)(n-m)}}\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{j+1}\left(\frac{1}{2^{\frac{3m-n}{2}}}+\frac{q}{2^{\frac{n+m}{2}}}\right)^{j-1}\right)^{\frac{1}{3}}$$

for $2 \le j \le t-1$, and

$$\beta := \frac{2^{tm}}{(t-1)!}\left(\frac{q^{t+1}}{t(t+1)\cdot 2^{t(n+1)}}\right)^{\frac{1}{2}},$$

and use Lemma 4.8. $\qquad\square$

The proof of Theorem 2 follows by taking $t = \lceil\frac{n+m}{n-m}\rceil$ in Lemma 4.9.

Proof of Theorem 2. The inequality (5) clearly holds for $q \ge \frac{1}{8}2^{\frac{m+n}{2}}$ since surely $Adv \le 1$ and

$$1 = 3\left(\frac{1}{8}\right)^{2/3} + 2\left(\frac{1}{8}\right) \le 3\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + 2\left(\frac{q}{2^{\frac{n+m}{2}}}\right) <$$

$$< 3\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{2/3} + 2\left(\frac{q}{2^{\frac{n+m}{2}}}\right) + 5\left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + \frac{1}{2}\left(\frac{2q}{2^{\frac{n+m}{2}}}\right)^{\frac{n}{n-m}}.$$

We therefore assume $q < \frac{1}{8}2^{\frac{m+n}{2}}$, and let $t := \lceil\frac{n+m}{n-m}\rceil$. Note that $5 \le m \le n-8$, since $\frac{n}{3} < m \le n - \log_2 n - 4$. Therefore,

$$\frac{2(t-1)}{2^m} \le \frac{2\cdot\frac{n+m}{n-m}}{2^m} = \frac{\frac{4m}{n-m}+2}{2^m} \le \frac{m+4}{2^{m+1}} \le \frac{9}{64},$$

$$\frac{2q}{2^n} = \frac{2}{2^{\frac{n-m}{2}}}\cdot\frac{q}{2^{\frac{n+m}{2}}} \le \frac{1}{64},$$

hence

$$4\left(1+\left(\frac{2(t-1)}{2^m}+\frac{2q}{2^n}\right)^{t-2}\right) \le 4\left(1+\left(\frac{9}{64}+\frac{1}{64}\right)\right) < 5. \tag{32}$$

Additionaly, using also that $t \ge 3$, since $m > \frac{n}{3}$,

$$2\sqrt[3]{2}\left(1+\frac{4}{2^m}\right)^{2/3} \le 2\sqrt[3]{2}\left(1+\frac{4}{2^5}\right)^{2/3} < 3, \tag{33}$$

$$2\sqrt[3]{100}\left(\frac{1}{2^{n-m}}\left(\frac{1}{2^{\frac{3m-n}{2}}}+\frac{q}{2^{\frac{n+m}{2}}}\right)\right)^{\frac{1}{3}} < 2\sqrt[3]{100}\left(\frac{1}{2^8}\left(1+\frac{1}{8}\right)\right)^{\frac{1}{3}} < 2, \tag{34}$$

$$\frac{\sqrt{2}}{\sqrt{t(t+1)}}\cdot\frac{1}{2^{\frac{n-m}{4}\left(t-\frac{n+m}{n-m}\right)}} \le \frac{\sqrt{2}}{\sqrt{3\cdot 4}} < \frac{1}{2}, \tag{35}$$

and finally, since $2q/2^{\frac{n+m}{2}} < 1$ and $\frac{t+1}{2} \ge \frac{n}{n+m}$,

$$\left(\frac{2q}{2^{\frac{n+m}{2}}}\right)^{\frac{t+1}{2}} < \left(\frac{2q}{2^{\frac{n+m}{2}}}\right)^{\frac{n}{n+m}}. \tag{36}$$

Since it is straightforward to verify that n, m, q, t satisfy all the conditions of Lemma 4.9, we get (5) by combining (31), (32), (33), (34), (35) and (36). $\qquad\square$

5 Discussion

We conclude with the following note. As mentioned above, the analysis in [2] is also based on examining the set S, but only for the particular choice of parameters: $t = 2$, $\alpha_1 = cq/2^{\frac{n-m+1}{2}}$, $\beta = 0$. Choosing α_1 to be proportional to the standard deviation of col_2 seems reasonable and natural (although in our analysis we employ a somewhat different choice). However, choosing $\beta = 0$ is artificial and too restrictive, and limiting t to be 2 is insufficient for getting the result for large m.

References

[1] S. Gilboa, S. Gueron, B. Morris, How many queries are needed to distinguish a truncated random permutation from a random function?, submitted, preprint available at `arXiv:1412.5204`.

[2] C. Hall, D. Wagner, J. Kelsey, B. Schneier, Building prfs from prps, in: Proceedings of CRYPTO-98: Advances in Cryptography, Springer Verlag, 1998, pp. 370-389.

[3] A. J. Stam, Distance between sampling with and without replacement, Statist. Neerlandica **32** (1978), no. 2, 81–91.